Statistical Properties of Loss Rate Estimators in Tree Topology

Weiping Zhu

Abstract

Three types of explicit estimators are proposed here to estimate the loss rates of the links in a network of the tree topology. All of them are derived by the maximum likelihood principle and proved to be either asymptotic unbiased or unbiased. In addition, a set of formulae are derived to compute the efficiencies and variances of the estimators that also cover some of the estimators proposed previously. The formulae unveil that the variance of the estimates obtained by a maximum likelihood estimator for the pass rate of the root link of a multicast tree is equal to the variance of the pass rate of the multicast tree divided by the pass rate of the tree connected to the root link. Using the formulae, we are able to evaluate the estimators proposed so far and select an estimator for a data set.

Index Terms

Correlation, Efficiency, Explicit Estimator, Loss Tomography, Maximum Likelihood, Variance.

I. INTRODUCTION

Network characteristics, such as link-level loss rate, delay distribution, available bandwidth, etc. are valuable information to network operations, development and researches. Therefore, a considerable attention has been given to network measurement, in particular to large networks that cross a number of autonomous systems, where security concerns, commercial interests, and administrative boundary make direct measurement impossible. To overcome the security and administrative obstacles, network tomography was proposed in [1], where the author suggests the use of end-to-end measurement and statistical inference to estimate the characteristics of interest. Since then, many works have been carried out to estimate various characteristics that cover loss tomography [2–11], delay tomography [12–16], loss pattern tomography [8], and so on. Despite the enthusiasm in loss tomography, there has been little work to study the statistical properties of an estimator with a finite sample size although some asymptotic properties are presented in the literature [2, 9]. The finite sample properties, such as efficiency and variance, differ from the asymptotic ones that are critical to the performance evaluation of an estimator since each of them unveil the quality and effectiveness of an estimator in a specific aspect. Apart from that, the finite sample properties can be used to select a better estimator, if not the best, from a group for a data set obtained from a specific circumstance. To fill the gap, we in this paper propose a number of maximum likelihood estimators (MLE) that can be solved explicitly for a network of the tree topology and provide the statistical properties of them. The statistical properties are further extended to cover the MLEs proposed previously. One of the most important discoveries is a set of formulae to compute the efficiency and variance of the estimates obtained by an estimator.

Weiping Zhu is with University of New South Wales, Australia, email w.zhu@adfa.edu.au

The approach proposed in [1] requires us to send probing packets, called probes, from some end-nodes called sources to the receivers located on the other side of the network, where the paths connecting the sources to the receivers cover the links of interest. To make the probes received informative in statistical inference, multicast or unicast-based multicast proposed in [5, 17] is used to send probes from a source to a number of receivers, via a number of intermediate nodes that replicate the arrived probes and then forward to its descendants. This process continues until either the probes reach the destinations or lost, which makes the observations of any two receivers correlated in some degree and the degrees vary depending on the interconnection between the receivers. Given the network topology used for sending probes and the observations obtained at receivers, we are able to create a likelihood function to connect the observation to the process described above. Since the number of correlations created by multicasting are proportional to the number of descendants attached to a node, the likelihood equation obtained for a node having many descendants is a high degree polynomial that requires an iterative procedure, such as the expectation and maximization (EM) or the Newton-Raphson algorithm, to approximate the solution. Using iterative procedure to solve a polynomial has been widely criticised for its computational complexity that increases with the number of descendants attached to the link or path to be estimated [5]. There has been a persistent effort in the research community to search for explicit estimators that are comparable in terms of accuracy to the estimators using iterative approach. To achieve this, we must have the statistical properties of the estimates obtained by an estimator, such as unbiasedness, efficiency, and variance. Unfortunately, there has been little work in a general form for the properties and the asymptotic properties obtained in [2, 9] has little use in this circumstance.

To overcome the problems stated above, we have undertaken a thorough and systematic investigation of the estimators proposed for loss tomography that aims at identifying the statistical principle and strategies that have been used or can be used in the tree topology. A number of findings are obtained in the investigation that show all of the estimators proposed previously rely on observed correlations to infer the loss/pass rates and most of them use all of the correlations available in estimation, such as the MLE proposed in [2]. However, the qualities of the correlations, measured by the fitness between a correlation and the corresponding observation, are very much ignored. Rather than using all of the correlations available in estimation, we propose here to use a small portion of high-quality ones and expect the estimates obtained by such an estimator are comparable to that considering all of the correlations. The investigation further leads to a number of findings that contribute to loss tomography in four-fold.

- A large number of explicit estimators are proposed on the basis of composite likelihood [18] that are divided into three groups: the block wised estimators (BWE), the reduce scaled estimators (RSE), and the individual based estimators (IBE).
- The estimators in BWE and IBE are proved to be unbiased and that in RSE are proved to be asymptotic unbiased as that proved in [9]. A set of formulae are derived for the efficiency and variances of the estimators in RSE and IBE, plus the MLE proposed in [2]. The formulae show the variance of the estimates obtained by a MLE can be exactly expressed by the pass rate of the path of interest and the pass rate of the subtrees connected to the path. The formulae also show the weakness of the result obtained in [9].
- The efficiency of the estimators in IBE are compared with each other on the basis of the Fisher information that shows an estimator considering a correlation involving a few observers can be more efficient than that considering more and the estimator proposed in [9] is the least efficient. A similar conclusion is obtained for the estimators in BWE.

- Using the formulae, we able to identify an efficient estimator by examining the end-to-end observation that makes model selection not only possible but also feasible. A number of simulations are conducted to verify this feature that also show the connection between efficiency and robustness of an estimator.

The rest of the paper is organised as follows. In Section II, we briefly introduce the previous works related to explicit loss rate estimators and point out the weakness of them. In Section III, we introduce the loss model, the notations, and the statistics used in this paper. Using the model and statistics, we derive a MLE that considers all available correlations for a network of the tree topology in Section IV. We then decompose the MLE into a number of components according to correlations and derive a number of likelihood equations for the components in Section V. A statistic analysis of the proposed estimators is presented in Section VI that details the statistical properties of the proposed estimators, one of them is the formulae to calculate the variances of various estimators. Simulation study is presented in Section VII that compares the performance of five estimators and shows the feasibility of selecting an estimator for a data set. Section VIII is devoted to concluding remark.

II. RELATED WORKS

Multicast Inference of Network Characters (MINC) is the pioneer of using the ideas proposed in [1] into practice, where a Bernoulli model is used to model the loss behaviors of a path. Using this model, the authors of [2] derive an estimator in the form of a polynomial that is one degree less than the number of descendants connected to the end node of the path of interest [2–4]. Apart from that, the authors obtain a number of results from asymptotic theory, such as the large number behaviour of the estimator and the dependency of the estimator variance on topology. Unfortunately, the results only hold if the sample size n grows indefinitely. In addition, if $n \to \infty$, almost all of the estimators proposed previously must have the same results and no one can tell the difference between them. In order to evaluate the performance of an estimator, experiments and simulation have been widely used but lead to little result since there are too many random factors affecting the results obtained from experiments and simulations.

To overcome the problem stated, simple and explicit estimators, such as that proposed in [9], are investigated that aims at reducing the complexity of an estimator and hopefully finding theoretical support for further development since a simple estimator may be easy to analyse. Using this strategy, the authors of [9] propose an explicit estimator that only considers a correlation, i.e. the correlation involving all descendants, and claim the same asymptotic variance for the estimates obtained by the estimator as that obtained by the estimator proposed in [2] to first order. The claim is obtained by applying the central limited theorem (CLT) on one of the results acquired by the asymptotic theory in [2], where the covariance between two descendants attached to the path of interest is obtained by assuming the loss rate of a link is very small and then the delta method is used to compute the asymptotic variance on the covariance matrix obtained by the asymptotic theory. The repeated use of the CLT makes the claim questionable and expensive to use in practice since the result only holds if $n \to \infty$. Apart from that, some sensitive parameters are cancelled out in the process. It is easy to prove that under the same condition, most of the estimators proposed so far can achieve the same result, if not better, as that proposed in [9].

In contrast to [9], [8, 19] propose an estimator that converts a general tree into a binary one and subsequently makes the likelihood equation into a quadratic equation of A_k that is solvable analytically. Experiments show the estimator preforms better than that in [9] since the estimator uses more information in estimation. Except experimental results, there is little statistical analysis to demonstrate why it is better than that proposed in [9] and how to improve from there. Although the author of [19]

proves the estimator is a MLE, there is a lack of other statistical properties, such as whether the MLE proposed in [19] is the same as that proposed in [2] and if not, how much difference between them.

To be able to evaluate the performance of an estimator, we need to have the statistical properties of the estimator, such as unbiasedness, efficiency, variance, and so on, that differ from the asymptotic ones by showing the quality of an estimator in a finite sample. To distinguish the properties from the asymptotic ones, we call them finite sample properties and there has been a lack of results for the finite sample properties. This paper aims to fill the gap and provides the properties.

III. ASSUMPTION, NOTATION AND SUFFICIENT STATISTICS

To make the following statistical analysis clear and rigorous, we need to use a large number of symbols that may overwhelm the readers who are not familiar with loss tomography. To assist them, the symbols will be gradually introduced through the paper, where the frequently used symbols will be introduced in the following two sections and the others will be brought up until needed. In addition, the most frequently used symbols and their meanings are presented in Table I for quick reference.

A. Assumption

We assume the probes multicasted from the source to receivers are independent and network traffic remains statistically stable during the measurement. In addition, the observation obtained at receivers is considered to be independent identical distributed ($i.i.d.$). Further, the losses occurred at a node or on a link are assumed to be $i.i.d$ as well.

B. Notation

As stated, a network of the tree topology is considered in this paper and denoted by $T = (V, E)$ that multicasts probes from the source to a number of receivers, where $V = \{v_0, v_1, ...v_m\}$ is a set of nodes and $E = \{e_1, ..., e_m\}$ is a set of directed links that connect the nodes in V. In addition, $v_k, k \in \{1, \cdots, m\}$ is often called node k and e_k called link k in the following discussion. By default, node 0 is the root node of the multicast tree to which the source is attached. Apart from being the root that does not have a parent, node 0 is different from others by having a single descendant, v_1, that is connected by e_1. Among the nodes in V, there are a number of them called leaf nodes that do not have any descendant but a receiver is attached to a leaf node. Because of this, we do not distinguish between a leaf node and a receiver and we use $R, R \subset V$ to denote them. Since there are m links to connect $m + 1$ nodes in T, the links and nodes are organised in such a way that if $f(i)$ is used to denote the parent of node i, e_i is the link connecting $v_{f(i)}$ to v_i. Figure 1 is an example of a multicast binary tree that is named and connected according to the rules.

A multicast tree, as a tree, can be decomposed into a number of multicast subtrees at node $i, i \in V \setminus (v_0 \vee R)$, where $T(i)$ denotes the multicast subtree that has $v_{f(i)}$ as its root, e_i as its root link, and $R(i)$ as the receivers attached to $T(i)$. In addition, we use d_i to denote the descendants attached to node i that is a nonempty set if $i \notin R$. If x is a set, $|x|$ is used to denote the number of elements in x. Thus, $|d_i|$ denotes the number of descendants in d_i. Using the symbols on Figure 1, we have $R = \{v_8, v_9, \cdots, v_{15}\}$, $R(v_2) = \{v_8, v_9, v_{10}, v_{11}\}$, $d_{v_2} = \{v_4, v_5\}$, and $|d_{v_2}| = 2$.

If n probes are sent from v_0 to R in an experiment, each of them gives rise of an independent realisation of the passing (loss) process X. Let $X^{(i)}, i = 1, ..., n$ donate the $i - th$ process, where $x_k^i = 1, k \in V$ if probe i reaches v_k; otherwise $x_k^i = 0$. The sample $Y = (x_j^{(i)})_{j \in R}^{i \in \{1,..,n\}}$ is the observation obtained in an experiment that can be divided into a number of

sections according to $R(k)$, where $Y_k, k \in V$ denotes the part of Y obtained by $R(k)$. In addition, each of the sections can be further divided into subsections $Y_x, x \subset d_k$ that is the part of observation obtained by $R(j), j \in x \wedge x \subset d_k$. Obviously, $Y_x \subset Y_k$. If we use y_j^i to denote the observation of receiver j for probe i, we have $y_j^i = 1$ if probe i is observed by receiver j; otherwise, $y_j^i = 0$.

Although loss tomography aims at estimate the loss rate of a link, the pass rate of the path connecting v_0 to $v_k, k \in V$ is often used as the parameter to be estimated. Let A_k be the pass rate of the path connecting v_0 to v_k that is defined as the percentage of the number of probes arrived at node k among the number of probes sent by the source. Given $A_k, k \in V \setminus v_0$, we are able to compute the pass rates of all links in E since there is a bijection from the pass rates of the paths to the pass rates of the links in a network of the tree topology. If α_k denotes the pass rate of link k we have

$$\alpha_k = \frac{A_k}{A_{f(k)}}. \tag{1}$$

Given α_k, we are able to compute the loss rate of link k that is equal to $\bar{\alpha}_k = 1 - \alpha_k$.

C. Statistics

To estimate A_k from Y, we need a likelihood function to connect the *i.i.d.* model defined previously to Y. To support the initiative of using a part of the available correlations to estimate A_k, a function, $n_k(x), x \subseteq d_k$, defined as follows is used to return the statistic for the likelihood function:

$$n_k(x) = \sum_{i=1}^{n} \bigvee_{\substack{j \in R(z) \\ z \in x}} y_j^i. \tag{2}$$

Obviously.

$$n_k(d_k) = \sum_{i=1}^{n} \bigvee_{j \in R(k)} y_j^i. \tag{3}$$

$n_k(x)$ is the number of probes, confirmed by the observation of $R(i), i \in x$, reaching node k. If N_k is used to denote the number of probes reaching node k, we have $N_k \geq n_k(d_k) \geq n_k(x), x \subset d_k$, where $n_k(d_k)$ and $n_k(x)$ are of statistics that can be used to estimate A_k.

To write a likelihood function of A_k with $n_k(d_k)$, β_k and γ_k are introduced to denote the pass rate of the subtrees rooted at node k and the pass rate of the special multicast tree that connects v_0 to node k and then to $R(k)$. Clearly $\gamma_k = A_k \cdot \beta_k, k \in V$ and $\hat{\gamma}_k = \frac{n_k(d_k)}{n}$ that is the empirical value of γ_k. Note that $\hat{\gamma}_j = \frac{n_j(j)}{n}, j \in R$ is the empirical pass rate of the path from the root to node j. Given the assumptions and definitions, the likelihood function of A_k for observation $n_k(d_k)$ is written as follows:

$$\mathcal{L}(A_k, n_k(d_k)) = (A_k \beta_k)^{n_k(d_k)} (1 - A_k \beta_k)^{n - n_k(d_k)}. \tag{4}$$

We can then prove $n_k(d_k)$ is a sufficient statistic with respect to (*wrt.*) the passing process of A_k for the observation obtained by $R(k)$. Rather than using the well known factorisation theorem in the proof, we directly use the mathematic definition of a sufficient statistic (See definition 7.18 in [20]) to achieve this. The definition *wrt.* the statistical model defined for the passing process is presented as a theorem here:

Theorem 1: Let $Y_k = \{X^{(1)}, ..., X^{(n)}\}$ be an i.i.d random sample, governed by $\mathcal{L}(A_k|Y_k)$. The statistic $n_k(d_k)$ is minimal sufficient for A_k in respect of the observation of Y_k.
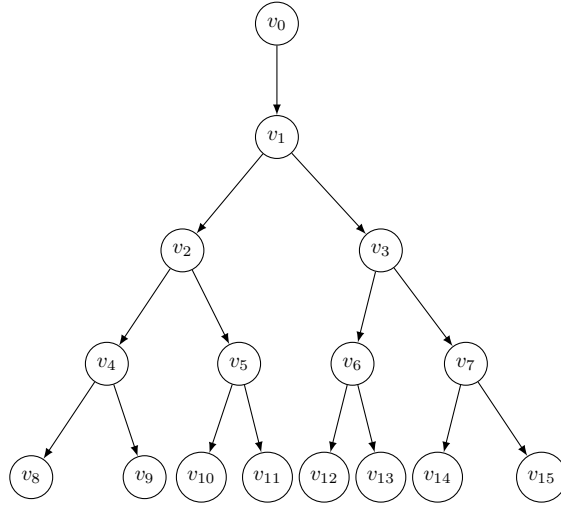
Fig. 1.   A Multicast Tree

Proof: According to the definition of sufficiency, we need to prove

$$\mathcal{L}(A_k|n_k(d_k) = t) = \frac{\mathcal{L}(A_k, n_k(d_k) = t)}{\mathcal{L}(n_k(d_k) = t)} \tag{5}$$

is independent of A_k.

Given (4), the passing process with observation of $n_k(d_k) = t$ is a random process that yields the binomial distribution as follows

$$\mathcal{L}(n_k(d_k) = t) = \binom{n}{t}(A_k\beta_k)^t(1 - A_k\beta_k)^{n-t}.$$

Then, we have

$$\begin{aligned}\mathcal{L}(A_k|n_k(d_k) = t) & = \frac{(A_k\beta_k)^t(1 - A_k\beta_k)^{n-t}}{\binom{n}{t}(A_k\beta_k)^t(1 - A_k\beta_k)^{n-t}.} \\ & = \frac{1}{\binom{n}{t}},\end{aligned} \tag{6}$$

which is independent of A_k. Then, $n_k(d_k)$ is a sufficient statistic.

Apart from the sufficiency, $n_k(d_k)$, as defined in (3), is a count of the probes reaching $R(k)$ that counts each probe once and once only regardless of how many receivers observe the probe. Therefore, $n_k(d_k)$ is a minimal sufficient statistic in regard to the observation of $R(k)$. ∎

D. Statistics considering a part of observation

Instead of using $n_k(d_k)$ to estimate A_k, we can use $n_k(x), x \subset d_k \wedge |x| \geq 2$, defined in Section III-C to estimate A_k. The difference between them is the number of correlations considered in estimation, where the latter is smaller than the former. As (4), $\beta_k(x), x \subset d_k$ is needed to express the pass rate of the subtrees consisting of $T(j), j \in x$. Given $n_k(x)$ and $\beta_k(x)$, we can also write a likelihood function of A_k and use the same procedure as that in Section III-C to prove $n_k(x)$ a sufficient statistic in the context of the observation obtained by $R(j), j \in x$. Further, an estimator on the observation of $R(j), j \in x$ can be created that will be discussed in Section V.

TABLE I

FREQUENTLY USED SYMBOLS AND DESCRIPTION

Symbol	Desciption
$T(k)$	the subtree rooted at link k.
d_k	the descendants attached to node k.
$R(k)$	the receivers attached to $T(k)$.
A_k	the pass rate of the path from v_0 to v_k.
β_k	the pass rate of the subtree rooted at node k.
$\beta_k(x)$	the pass rate of the subtree consisting of $T(j), j \in x \wedge x \subset d_k$.
γ_k	$A_k * \beta_k$, pass rate from v_0 to $R(k)$.
N_k	the number of probes reaching node k.
x_k^i	the state of v_k for probe i.
$\sum_k$	the σ-algebra created from d_k.
n	the number of probes sent in an experiment,
$n_k(d_k)$	the number of probes reaches $R(k)$.
$n_k(x)$	the number of probes reaches the receivers attached to $T(j), j \in x$.
$I_k(x)$	the number of probes observed by the members of x.
Y	the observation obtained in an experiment.
$Y_k, k \in V$	the part of Y obtained by $R(k)$.
$Y_x, x \subset d_k$	the part of Y obtained by $R(j), j \in x$.

IV. ESTIMATOR ANALYSIS

This section is dedicate to the analysis of the MLE that considers all of the correlations available in observation. By the analysis, we are able to identify all of the correlations in observation and find the connections among them that will set up the foundation for various explicit estimators.

A. Maximum Likelihood Estimator based on $n_k(d_k)$

Turning the likelihood function presented in (4) into a log-likelihood function, we have

$$\log \mathcal{L}(A_k|Y_k) = n_k(d_k) \log(A_k \beta_k) + (n - n_k(d_k)) \log(1 - A_k \beta_k). \tag{7}$$

Differentiating (7) *wrt.* A_k and letting the derivatives be 0, we have

$$\frac{n_k(d_k)}{A_k} - \frac{(n - n_k(d_k))\beta_k}{1 - A_k \beta_k} = 0, \tag{8}$$

and then

$$A_k \beta_k = \frac{n_k(d_k)}{n}. \tag{9}$$

Since neither A_k nor β_k can be solved from (9), we need to consider other correlations and then derive a MLE. Given the *i.i.d.* model assumed previously and the multicast used in probing, we have the following equation to link the observation of $R(k)$ to β_k

$$1 - \beta_k = \prod_{j \in d_k} (1 - \frac{\gamma_j}{A_k}). \tag{10}$$

Solving β_k from (10) and using it in (8), we have a MLE as

$$1 - \frac{n_k(d_k)}{n \cdot A_k} = \prod_{j \in d_k} (1 - \frac{\gamma_j}{A_k}). \tag{11}$$

Using γ_k to replace $\dfrac{n_k(d_k)}{n}$ since the latter is the empirical value of the former, we have a likelihood equation as follows:

$$1 - \frac{\gamma_k}{A_k} = \prod_{j \in d_k} (1 - \frac{\gamma_j}{A_k}) \tag{12}$$

that is identical to the estimator proposed in [2].

B. Predictor and Observation

To make the correlations involved in (11) visible, we expand the left hand side (LHS) and the right hand side (RHS) of (11), where the terms obtained from the LHS are called observations and the terms from the RHS are called correlations. The correlation is also called the *predictor* since it predicates the observation received in an experiment. For instance, $\gamma_i \cdot \gamma_j / A_k, i, j \in d_k \wedge i \neq j$ is the predictor of the probes simultaneously observed by the receivers attached to subtree i and subtree j, i.e. there is at least one receiver from each subtree.

To represent the correlations involved in (11), a σ-algebra, S_k, is created over d_k and let $\Sigma_k = S_k \setminus \emptyset$ be the non-empty sets in S_k. Each member in Σ_k corresponds to a pair of a predictor and its observation. If the number of elements in a member of Σ_k is defined as the degree of the correlation, Σ_k can be divided into $|d_k|$ exclusive groups, one for a degree of correlations that vary from 1 degree to $|d_k|$ degree. Let $S_k(i), i \in \{1, \cdots, |d_k|\}$ denote the group that considers i degree correlations. For example, if $d_k = \{i, j, k, l\}$, $S_k(2) = \{(i, j), (i, k), (i, l), (j, k), (j, l), (k, l)\}$ consists of the pairwise correlations in d_k, and $S_k(3) = \{(i, j, k), (i, j, l), (i, k, l), (j, k, l)\}$ contains all of the triplet-wise correlations.

Given Σ_k, $n_k(d_k)$ can be decomposed into the probes that are observed simultaneously by the members of Σ_k that is defined as if $x \in \Sigma_k$ and $|x| > 1$, a probe observed by x if and only if at least a receiver attached to subtree $j, j \in x$ observes the probe. We call such an observation simultaneous observation. To explicitly express $n_k(d_k)$ by $n_j(d_j), j \in d_k$, $I_k(x), x \in \Sigma_k$ is introduced to return the number of probes observed simultaneously by x in an experiment. Let u_j^i be the observation of $R(j)$ for probe i that is defined as:

$$u_j^i = \bigvee_{k \in R(j)} y_k^i,$$

then

$$I_k(x) = \sum_{i=1}^{n} \bigwedge_{j \in x} u_j^i, \ x \in \Sigma_k. \tag{13}$$

If $x = (j)$,

$$I_k(x) = n_j(d_j), j \in d_k,$$

Given the above, $n_k(d_k)$ can be decomposed as:

$$n_k(d_k) = \sum_{i=1}^{|d_k|} (-1)^{i-1} \sum_{x \in S_k(i)} I_k(x). \tag{14}$$

(14) states that $n_k(d_k)$ is equal to a series of $I_k(x), x \in S_k(i)$ that are overlapped each other. To ensure each probe observed by $R(k)$ is counted once and once only in $n_k(d_k)$, we need to use the alternating adding and subtracting operations to eliminate duplication.

C. Correspondence between Predictors and Observations

Given (14), we are able to prove the MLE proposed in [2] considers all of the correlations in Σ_k and have the following theorem.

Theorem 2: 1) (11) is a full likelihood estimator that considers all of the correlations in Σ_k;

2) (11) consists of observed values and their predictors, one for a member of Σ_k; and

3) the estimate obtained from (11) is a fit that minimises an alternating differences between observed values and corresponding predictors.

Proof: (11) is a full likelihood estimator that considers all of the correlations in d_k. To prove 2) and 3), we expand the both sides of (11) to pair the observed values with the predictors of them according to S_k. We take three steps to achieve the goal.

1) If we use (14) to replace $n_k(d_k)$ from LHS of (11), the LHS becomes:

$$1 - \frac{n_k(d_k)}{n \cdot A_k} = 1 - \frac{1}{n \cdot A_k}\Big[\sum_{i=1}^{|d_k|}(-1)^{i-1}\sum_{x \in S_k(i)} I_k(x)\Big]. \tag{15}$$

2) If we expand the product term located on the RHS of (11), we have:

$$\prod_{j \in d_k}\Big(1 - \frac{\gamma_j}{A_k}\Big) = 1 - \sum_{i=1}^{|d_k|}(-1)^{i-1}\sum_{x \in S_k(i)}\frac{\prod_{j \in x}\gamma_j}{A_k^i} \tag{16}$$

where the alternative adding and subtracting operations intend to remove the impact of redundant observation.

3) Deducting 1 from both (15) and (16) and then multiplying the results by A_k, (11) turns to

$$\sum_{i=1}^{|d_k|}(-1)^i\sum_{x \in S_k(i)}\frac{I_k(x)}{n} = \sum_{i=1}^{|d_k|}(-1)^i\sum_{x \in S_k(i)}\frac{\prod_{j \in x}\gamma_j}{A_k^{i-1}}. \tag{17}$$

It is clear there is a correspondence between the terms across the equal sign, where the terms on the LHS are the observed values and the terms on the RHS are the predictors. If we rewrite (17) as

$$\sum_{i=1}^{|d_k|}(-1)^i\sum_{x \in S_k(i)}\Big(\frac{I_k(x)}{n} - \frac{\prod_{j \in x}\gamma_j}{A_k^{i-1}}\Big) = 0, \tag{18}$$

the correspondence between correlations and observed values becomes obvious. ∎

To distinguish the MLE from those proposed in this paper, we call it original MLE in the rest of the paper.

V. Explicit Estimators based on Composite Likelihood

(18) shows that the original MLE takes into account all of the correlations in Σ_k. If the number of subtrees rooted at node k is larger than 6, the estimator is a high degree polynomial that could not be solved analytically. To have an explicit estimator in those circumstances, we need to reduce the number of correlations considered in estimation and there are a number of strategies to achieve this. We here propose three of them and use composite likelihood that is also called pseudo-likelihood by Besag in [21] to create likelihood functions for the strategies. The three strategies are named reduce scaled, block-wised, and individual based, respectively. The reduce scaled strategy, as named, is a small version of the original MLE that selectively removes a number of subtrees rooted at node k from consideration and then uses the maximum likelihood principle on the rest to estimate A_k. The block-wised strategy differs from the reduce scaled one by dividing all available correlations considered

by the original MLE into a number of blocks, one for a degree of correlations, from pairwise to d_k-wise. The individual based one, in contrast to the other two, considers a correlation at a time that leads to a large number of estimators.

A. Reduce Scaled Estimator (RSE)

Rather than considering all of the correlations in Σ_k, the correlations can be divided into groups according to the subtrees rooted at node k. Let $x, x \subset d_k$ be the group to be considered by an estimator in RSE. The log-likelihood function considering the correlations x is as follows:

$$\log L(A_k|Y_x) = n_k(x)\log(A_k\beta_k(x)) + (n - n_k(x))\log(1 - A_k\beta_k(x)) \tag{19}$$

where $n_k(x)$ as defined in III-D is the number of probes reaching node k confirmed from the observations of the receivers attached to $T(j), j \in x$ and $\beta_k(x)$ is the pass rate of $T(j), j \in x$ that can be expressed as

$$1 - \beta_k(x) = \prod_{j \in x}(1 - \frac{\gamma_j}{A_k}). \tag{20}$$

Then, a similar likelihood equation as (11) is obtained and presented as follows:

$$1 - \frac{n_k(x)}{n \cdot A_k} = \prod_{j \in x}(1 - \frac{\gamma_j}{A_k}). \tag{21}$$

If $|x| < 5$, the equation is solvable analytically. The estimators in RSE are denoted by $Am_k(x), x \subset d_k$.

B. Block-wised Estimator (BWE)

(18) shows that the correlations involved in the original MLE can be divided into $|d_k| - 1$ blocks, from pairwise to $|d_k|$-wise. Each of them can be written as a likelihood function. In order to use a unique likelihood function for all of them, we let the likelihood function considering single correlation be 1. Then, the i-wise likelihood function denoted as $L_c(i; A_k; y)$ can be expressed uniformly.

Definition 1: There are a number of composite likelihood functions, one for a degree of correlations, varying from pairwise to $|d_k|$-wise. The composite likelihood function $L_c(i; A_k; y), i \in \{2, \cdots, |d_k|\}$ has a form as follows:

$$
\begin{aligned}
L_c(i; A_k; y) &= \frac{\prod_{x \in S(i)}(A_k\beta_k(x))^{n_k(x)}(1 - A_k\beta_k(x))^{n-n_k(x)}}{\prod_{x' \in S(i-1)}(A_k\beta_k(x'))^{n_k(x')}(1 - A_k\beta_k(x'))^{n-n_k(x')}}. \\
& i \in \{2, \cdots, |d_k|\}
\end{aligned}
\tag{22}
$$

Let $A_k(i)$ be the estimator derived from $L_c(i; A_k; y)$. Then, we have the following theorem.

Theorem 3: Each of the composite likelihood equations obtained from (22) is an explicit estimator of A_k that is as follows:

$$A_k(i) = \Big(\frac{\sum_{x \in S_k(i)} \prod_{j \in x} \gamma_j}{\sum_{x \in S_k(i)} \frac{I_k(x)}{n}}\Big)^{\frac{1}{i-1}}, i \in \{2, .., |d_k|\}. \tag{23}$$

Proof: Firstly, we can write (22) into a log-likelihood function and differentiate the log-likelihood function *wrt* A_k. As (9), we cannot solve A_k or $\beta_k(x)$ directly from the derivative and we need to consider other correlations as (10). We then have an equation as

$$\frac{\partial \log L_c(i, A_k; y)}{\partial A_k} = \sum_{x \in S(i)}\Big[1 - \frac{\gamma_k(x)}{A_k} - \prod_{q \in x}(1 - \frac{\gamma_q}{A_k})\Big] - \sum_{x' \in S(i-1)}\Big[1 - \frac{\gamma_k(x')}{A_k} - \prod_{q \in x'}(1 - \frac{\gamma_q}{A_k})\Big] \tag{24}$$

The two summations can be expanded as (11) and only the terms related to the i-wise correlation left since all other terms in the first summation are canceled by the terms of the second summation. The likelihood equation as (23) follows. ∎

In the rest of the paper, $A_k(i)$ is used to refer to the $i-wise$ estimator and $\widehat{A}_k(i)$ refers to the estimate obtained by $A_k(i)$.

C. Individual based Estimator (IBE)

Instead of considering a block of correlations together, we can consider a correlation at at time and have a large number of estimators. Each of them has a similar likelihood function as (19), where $\beta_k(x)$ and $n_k(x)$ are replaced by $\psi_k(x)$ and $I_k(x)$, respectively. $\psi_k(x) = \prod_{j \in x} \alpha_j \beta_j, x \subseteq d_k$, is the pass rate of $T(j), j \in x$. If $\Sigma'_k = \Sigma_k \setminus S_k(1)$ is the correlations considered by IBE, the log-likelihood function for A_k given observation $I_k(x)$ is equal to

$$L(A_k|I_k(x)) = I_k(x) \log(A_k \psi_k(x)) + (n - I_k(x)) \log(1 - A_k \psi_k(x)), \ x \in \Sigma'_k. \tag{25}$$

We then have the following theorem.

Theorem 4: Given (25), $A_k \psi_k(x)$ is a Bernoulli process. The MLE for A_k given $I_k(x)$ equals to

$$Al_k(x) = \Big(\frac{\prod_{j \in x} \gamma_j}{\frac{I_k(x)}{n}} \Big)^{\frac{1}{|x|-1}}. \ x \in \Sigma'_k \tag{26}$$

Proof: Using the same procedure as that used in IV-A, we have the theorem. ∎

Comparing (23) with (26), we can find that $\widehat{Al}_k(x)$, where $|x| = i$, is a type of geometric mean and $\widehat{A}_k(i)$ is the arithmetic mean of $\widehat{Al}_k(x), x \in S_k(i)$. Therefore, $A_k(i)$ is more robust than $Al_k(x)$.

Using and combining the strategies presented here, we can have various explicit estimators that cover those proposed previously. For instance, the estimator proposed in [8, 19] is one of them that divides d_k into two groups and only considers the pairwise correlations between the members of the two groups. Therefore, although the estimator proposed in [8, 19] is a MLE in terms of the observation used in estimation, it is not the same as (12).

VI. PROPERTIES OF THE ESTIMATORS

It is known that if a MLE is a function of the sufficient statistic, it is asymptotically unbiased, consistent and asymptotically efficient. Thus, the original MLE and all of the estimators proposed in this paper have the properties. Apart from them, we are interested in whether some of the estimators have better properties than them, such as, unbiasedness, uniqueness, variance, and efficiency, that can be used to evaluate the estimators. This section is devoted to present them that consist of a number of theorems and corollaries.

A. Unbiasedness and Uniqueness of $Al(x)$ and $A_k(i)$

This subsection is focused on the unbiasedness of the estimators in IBE and BWE although the statistic used by the latter is not minimal sufficient. For $Al_k(x), x \in \Sigma'_k$, we have the following theorem.

Theorem 5: $Al_k(x)$ is a unbiased estimator.

Proof: Let $z_j, j \in d_k$ be the pass rate of $T(j)$ and let $\overline{A_k} = \frac{N_k}{n}$ be the sample mean of A_k. Note that z_j and $z_l, j, l \in d_k$ are independent from each other if $j \neq l$. In addition, $z_j, j \in d_k$ is independent from A_k. Because of this, $x_k^i \prod_{j \in x} z_j$ is used to replace $\bigwedge_{j \in x} y_j^i$ in the following derivation since the latter is equal to $\prod_{j \in x} y_j^i$ that is equal to $x_k^i \prod_{j \in x} z_j$. We then have

$$E(\widehat{Al}_k(x)) = E\left(\left(\frac{\prod_{j\in x}\hat{\gamma}_j}{\frac{I_k(x)}{n}}\right)^{\frac{1}{|x|-1}}\right)$$

$$= E\left(\left(\frac{\prod_{j\in x}\frac{n_j(d_j)}{n}}{\frac{\sum_{i=1}^{n}\bigwedge_{j\in x}y_j^i}{n}}\right)^{\frac{1}{|x|-1}}\right)$$

$$= E\left(\left(\frac{(\frac{N_k}{n})^{|x|}\prod_{j\in x}\frac{n_j(d_j)}{N_k}}{\frac{N_k}{n}\frac{\sum_{i=1}^{N_k}\prod_{j\in x}z_j}{N_k}}\right)^{\frac{1}{|x|-1}}\right)$$

$$= E\left(\frac{N_k}{n}\right)E\left(\left(\frac{\prod_{j\in x}\frac{1}{N_k}\sum_{i=1}^{N_k}z_j}{\sum_{i=1}^{N_k}\frac{1}{N_k}\prod_{j\in x}z_j}\right)^{\frac{1}{|x|-1}}\right)$$

$$= E\left(\overline{A_k}\right) \quad\quad (27)$$

The theorem follows. ∎

Given theorem 5, we have the follow corollary.

Corollary 1: $A_k(i)$ is a unbiased estimator.

Proof: According to theorem 5, we have

$$E(\widehat{A}_k(i)) = E\left(\overline{A_k}\right)E\left(\left(\frac{\sum_{x\in S(i)}\prod_{j\in x}\frac{1}{N_k}\sum_{i=1}^{N_k}z_j}{\sum_{x\in S(i)}\sum_{i=1}^{N_k}\frac{1}{N_k}\prod_{j\in x}z_j}\right)^{\frac{1}{i-1}}\right)$$

$$= E\left(\overline{A_k}\right) \quad\quad (28)$$

∎

Given $Al_k(x), x \in \Sigma'_k$ and $A_k(i)$ are unbiased estimators, we can prove the uniqueness of $A_k(i)$.

Theorem 6: If

$$\sum_{x\in S_k(i)}\prod_{j\in x}\hat{\gamma}_j < \sum_{x\in S_k(i)}\frac{I_k(x)}{n},$$

there is only one solution in $(0,1)$ for $\widehat{A}_k(i), 2 \leq i \leq |d_k|$.

Proof: Since the support of A_k is in (0,1), we can reach this conclusion from (23). ∎

B. Efficiency of $Al_k(x)$, $Am_k(x)$, and the original MLE

Apart from asymptotically efficiency stated previously for the MLEs using sufficient statistics, we are interested in the efficiency of the estimators proposed in this paper. Given (25), we have the following theorem for the Fisher information of an observation, y, for the estimators in IBE, i.e. $Al_k(x), x \in \Sigma'_k$.

Theorem 7: The Fisher information of y on $Al_k(x), x \subset d_k$ is equal to $\frac{\psi_k(x)}{A_k(1-A_k\psi_k(x))}$.

Proof: Considering $I_k(x) = y$ is the observation of the receivers attached to x, we have the following as the likelihood function of the observation:

$$L(A_k|y) = y\log(A_k\psi_k(x)) + (1-y)\log(1-A_k\psi_k(x)). \quad\quad (29)$$

Differentiating (29) *wrt* A_k, we have

$$\frac{\partial L(A_k|y)}{\partial A_k} = \frac{y}{A_k} - \frac{(1-y)\psi_k(x)}{1-A_k\psi_k(x)} \qu\quad (30)$$

We then have

$$\frac{\partial^2 L(A_k|y)}{\partial A_k^2} \quad = \quad -\frac{y}{A_k^2} - \frac{(1-y)\psi_k(x)^2}{(1-A_k\psi_k(x))^2} \tag{31}$$

If $\mathcal{I}(Al_k(x)|y)$ is used to denote the Fisher information of observation y for A_k in $Al_k(x)$, we have

$$\begin{aligned} \mathcal{I}(Al_k(x)|y) \quad &= \quad -E\left(\frac{\partial^2 L(A_k|y)}{\partial A_k^2}\right) \\ &= \quad \frac{E(y)}{A_k^2} + \frac{E(1-y)\psi_k(x)^2}{(1-A_k\psi_k(x))^2} \\ &= \quad \frac{\psi_k(x)}{A_k(1-A_k\psi_k(x))} \end{aligned} \tag{32}$$

that is the information provided by y for A_k. $\blacksquare$

Given (32), we are able to have a formula for the Fisher information of the original MLE and the estimators in RSE. In order to achieve this, let $\beta_k(d_k) = \beta_k$. Then, we have the following corollary.

Corollary 2: The Fisher information of observation y for A_k in the original MLE and $Am_k(x), x \subseteq d_k$ is equal to

$$\frac{\beta_k(x)}{A_k(1-A_k\beta_k(x))}, \ x \subseteq d_k. \tag{33}$$

Proof: Replacing $n_k(d_k)$ or $n_k(x)$ by y and replacing $n - n_k(d_k)$ or $n - n_k(x)$ by $1 - y$ from (7) and (19), respectively, and then using the same procedure as that used in the proof of theorem 7, the corollary follows. $\blacksquare$

Because of the similarity between (32) and (33), the two equations have the same features in terms of support, singularity, and maximums. After eliminating the singular points, the support of A_k is in $(0, 1)$ and the support of $\beta_k(x)$ (or $\psi_k(x)$) is in $[0, 1]$. Both (32) and (33) are convex functions in the support and reach the maximum at the points of $A_k \to 1, \beta_k(x) = 1$ (or ($\psi_k(x) = 1$) and $A_k \to 0, \beta_k(x) = 1$ (or ($\psi_k(x) = 1$). Given A_k, (33) is a monotonic increase function of $\beta_k(x)$ whereas (32) is a monotonic increase function of $\psi_k(x)$.

Despite the similarity between (32) and (33), $Am_k(x)$ and $Al_k(x)$ react differently if x is replaced by $y, x \subset y$ in terms of efficiency that leads to two corollaries, one for each of them.

Corollary 3: $Am_k(y)$ is more efficient than $Am_k(x)$ if $x \subset y$.

Proof: If $x \subset y$, $\beta_k(x) \leq \beta_k(y)$ and then we have the corollary. $\blacksquare$

For $Al_k(x)$, we have

Corollary 4: The efficiency of $Al_k(x), x \in \Sigma'_k$ forms a partial order that is identical to that formed on the inclusion of the members in Σ'_k, where the most efficient estimator must be one of the $Al_k(x), x \in S_k(2)$ and the least efficient one must be $Al_k(d_k)$.

Proof: According to Theorem 7, the efficiency of $Al_k(x)$ is determined by $\psi_k(x)$, where $\psi_k(x) = \prod_{j \in x} \alpha_j \beta_j$. If $x \subset y$, we have

$$\begin{aligned} \psi_k(y) \quad &= \quad \prod_{j \in y} \alpha_j \beta_j \\ &= \quad \psi_k(x) \prod_{j \in (y \setminus x)} \alpha_j \beta_j \\ &< \quad \psi_k(x) \end{aligned} \tag{34}$$

. Therefore, the order of the efficiency of $Al_k(x), x \in \Sigma'_k$ shares that of the inclusion in Σ'_k, where $x, x \in S_k(2)$ are the members of Σ'_k that have the minimal number of elements. In contrast to $\psi_k(x), x \in S_k(2)$, $\psi_k(d_k) \leq \psi_k(x)$ since $\forall x, x \in \Sigma'_k, x \subseteq d_k$. Then, the corollary follows. ∎

C. Variance of $Al_k(x)$, $Am_k(x)$, and the original MLE

The estimator specified by (12), $Am_k(x)$, and $Al_k(x)$ are of MLEs that have different focuses on the observations obtained. Despite the difference between them, they share a number of features, including likelihood function and efficient equation. In addition, the variances of them are expressed by a general function showing the connection between A_k and the pass rate of the subtree considered in estimation. Let mle denote all of them. Then, we have a theorem for the variances of the estimators in mle.

Theorem 8: The variances of the estimators in mle equal to

$$var(mle) = \frac{A_k(1 - A_k\delta_k(x))}{\delta_k(x)}, \ x \subseteq d_k \tag{35}$$

where $\delta_k(x)$

$$\delta_k(x) = \begin{cases} \beta_k(x), & \text{For original MLE and } Am_k(x); \\ \psi_k(x), & \text{For } Al_k(x). \end{cases}$$

Proof: The passing process described by (25) is a Bernoulli process that falls into the exponential family and satisfies the regularity conditions presented in [22]. Thus, the variance of an estimator in mle reaches the Cramér-Rao bound that is the reciprocal of the Fisher information. ∎

(35) can be written as

$$\frac{A_k}{\delta_k(x)} - A_k^2 \tag{36}$$

which shows:

1) the estimates obtained by an estimator spread out more widely than that obtained by direct measurement. The wideness is determined by $\delta_k(x)$, the pass rate of the subtrees connecting node k to observers. If $\delta_k(x) = 1$, there is no further spread-out than that obtained by direct measurement. Otherwise, the variance increases as the decreases of δ_k and in a super linear fashion.

2) the variance of the estimates obtained by an estimator is monotonically increasing as the depth of the subtree rooted at node k since the pass rate of a subtree decreases as its depth, i.e., the pass rate of an i-level tree, say A, is larger than that of the $i + 1$-level one that is extended from the i-level one;

3) the variance of the estimates obtained by an estimator in mle is a monotonically decreasing function of $\delta_k(x)$.

The three points confirm some of the experiment results reported previously, such as the dependency of variance on topology reported in [2]. Note than despite 3), the variance of $Am_k(x)$ can be the same as that of $Am_k(y), x \subset y$ if $\beta_k(x) = \beta_k(y)$. So does $Al_k(x)$. In other words, if the probes observed by $R(j), j \in (y \setminus x)$ are included in that observed by $R(i), i \in x$, the estimate obtained by $Am_k(x)$ is the same as that obtained by $Am_k(y)$.

D. Efficiency and Variance of BWE

As stated, the estimate obtained by $A_k(i)$ is a type of the arithmetic mean of $Al_k(x), x \in S_k(i)$ that has the same advantages and disadvantages as the arithmetic mean. Thus, $A_k(i)$ is more robust and efficient than that of $Al_k(x), x \in S(i)$ since the former considers more probes than the latter in estimation although some of the probes may be considered more than once. Because of this, (32) cannot be used to evaluate the efficiency of an estimator in BWE. Despite this, we can put a range for the information obtained by $A_k(i)$ that is

$$\frac{\psi_k(x)}{A_k(1 - A_k\psi_k(x))} \leq \mathcal{I}(A_k(i)|y) \leq \binom{d_k}{i} \frac{\psi_k(x)}{A_k(1 - A_k\psi_k(x))}. \ |x| = i. \tag{37}$$

In addition, $A(i)$ is at least as efficient as $A(i + 1)$ and the variance of $A(i)$ is at least as small as that of $A(i + 1)$ since $\sum_{x \in S_k(i)} I_k(x) \leq \sum_{x \in S_k(i+1)} I_k(x)$.

E. Example

We use an example to conclude this section that illustrate the differences of the variances obtained from the estimates of four estimators. The four estimators are: direct measurement, the original MLE, $Al_k(x), |x| = 2$ and $Al_k(d_k)$, respectively. The setting used here is identical to that presented in [9], where node k has three children with a pass rate of $\alpha, 0 < \alpha \leq 1$, and the pass rate from the root to node k is also equal to α. Using (35), we have the variances of them that are presented below:

1) $\alpha - \alpha^2$,
2) $\frac{1}{3(1-\alpha)+\alpha^2} - \alpha^2$,
3) $\frac{1}{\alpha} - \alpha^2$, and
4) $\frac{1}{\alpha^2} - \alpha^2$.

The difference between them becomes obvious as α decreases from 1 to 0.99, where the variances of the four estimators change from 0 to 0.01, 0.01, 0.03, and 0.04, respectively. The variance of $Al_k(d_k)$ is 4 times of that of the original MLE that is significantly different from that obtained in [9]. Although the variances are decreased as the number of probes multicasted, the ratio between them remains.

VII. MODEL SELECTION AND SIMULATION

The large number of estimators in IBE, RSE and BWE, plus the original MLE, make model selection possible. However, to find the most suitable one in terms of efficiency and computational complexity is a hard task since the two goals conflict each other. Although one is able to identify the the most suitable estimator by computing the Kullback-Leigh divergence or the composite Kullback-Leigh divergence of the estimators, the cost of computing the Akaike information criterion (AIC) for each of the estimators makes this approach prohibitive. Nevertheless, the derivation of (33) successfully solves the problem in some degree since (33) shows the most suitable estimator should have a bigger $\beta_k(x)$ which can be obtained from end-to-end observation since $\beta_k(x) \propto \prod_{j \in x} \gamma_j$.

Estimators	OMLE		$A_k(2)$		$A_k(3)$		$Al_k(x), |x| = 2$		$Al_k(x), |x| = 3$	
samples	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var
300	0.0088	1.59E-05	0.0088	1.59E-05	0.0088	1.64E-05	0.0087	1.59E-05	0.0087	1.61E-05
900	0.0092	7.76E-06	0.0092	7.82E-06	0.0091	7.84E-06	0.0092	7.90E-06	0.0092	8.15E-06
1500	0.0096	4.55E-06	0.0096	4.55E-06	0.0096	4.80E-06	0.0096	4.78E-06	0.0096	4.33E-06
2100	0.0097	3.14E-06	0.0097	3.11E-06	0.0097	3.14E-06	0.0097	3.02E-06	0.0097	3.08E-06
2700	0.0100	1.72E-06	0.0100	1.72E-06	0.0100	1.74E-06	0.0100	1.81E-06	0.0100	1.83E-06

TABLE II

SIMULATION RESULT OF A 8-DESCENDANT TREE WITH LOSS RATE=1%

Estimators	OMLE		$A_k(2)$		$A_k(3)$		$Al_k(x), |x| = 2$		$Al_k(x), |x| = 3$	
samples	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var
300	0.0088	1.59E-05	0.0089	1.64E-05	0.0089	1.68E-05	0.0091	2.36E-05	0.0088	1.95E-05
900	0.0091	7.76E-06	0.0091	7.80E-06	0.0091	7.83E-06	0.0092	9.74E-06	0.0091	8.67E-06
1500	0.0096	4.55E-06	0.0096	4.72E-06	0.0096	4.81E-06	0.0097	4.36E-06	0.0096	4.45E-06
2100	0.0097	3.14E-06	0.0097	3.11E-06	0.0097	3.11E-06	0.0098	3.39E-06	0.0097	3.04E-06
2700	0.0100	1.72E-06	0.0100	1.69E-06	0.0100	1.67E-06	0.0101	2.11E-06	0.0100	1.90E-06

TABLE III

SIMULATION RESULT OF A 8-DESCENDANT TREE, 6 OF THE 8 HAVE LOSS RATE=1% AND THE OTHER 2 HAVE LOSS RATE=5%

A. Simulation

To compare the effectiveness, robustness, and sensitivity of the estimators between the original MLE, $A_k(i)$, and $Al_k(x)$, three rounds of simulations are conducted in three settings. The multicast tree used in the simulations having a path/link from the root to node k that has 8 subtrees connecting to the receivers. Five estimators: the original MLE (OMLE), $A_k(2)$, $A_k(3)$, $Al_k(x), |x| = 2$, and $Al_k(x), |x| = 3$, are compared against each other in the simulation. The number of samples used in the simulations varies from 300 to 2700 in a step of 600. For each sample size, 20 experiments with different initial seeds are carried out and the means and variances of the estimates obtained by the five estimators are presented in three tables, from Table II to Table IV.

Table II is the results obtained in the first round that sets the loss rate of the subtrees to 1%, so does the loss rate of the path from the root to node k. The result shows when the sample size is small, the estimates obtained by all estimators are drifted away from the true value that indicates the data obtained is not enough. With the increase of sample size, the estimates gradually approach to the true value and all of the estimators achieve a similar outcome. As expected, the variances decrease

Estimators	OMLE		$A_k(2)$		$A_k(3)$		$Al_k(x), |x| = 2$		$Al_k(x), |x| = 3$	
samples	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var
300	0.0503	2.15E-04	0.0504	2.15E-04	0.0505	2.14E-04	0.0508	2.18E-04	0.0505	2.16E-04
900	0.0511	5.85E-05	0.0511	5.81E-05	0.0511	5.79E-05	0.0512	5.79E-05	0.0512	5.88E-05
1500	0.0502	2.24E-05	0.0502	2.24E-05	0.0502	2.23E-05	0.0503	2.33E-05	0.0502	2.32E-05
2100	0.0507	1.16E-05	0.0507	1.19E-05	0.0507	1.20E-05	0.0507	1.09E-05	0.0507	1.13E-05
2700	0.0507	1.31E-05	0.0507	1.34E-05	0.0507	1.35E-05	0.0508	1.35E-05	0.0507	1.34E-05

TABLE IV

SIMULATION RESULT OF A 8-DESCENDANT TREE, THE LOSS RATE OF THE ROOT LINK=5%, 4 OF THE 8 HAVE LOSS RATE=1% AND THE OTHER 4 HAVE LOSS RATE=5%

as the sample size that is agreed with (35) and there is no significant difference among the estimators since all of the subtrees connected to node k have the same loss rates. Despite this, the variance of $Al_k(x), |x| = 2$ is slightly better than that of $Al_k(x), |x| = 3$ as specified by Theorem 8.

To compare the sensitivity and robustness, another round simulation is carried out on the same network. The difference between this round and the previous one is the loss rates of the subtrees connected to node k, where 6 of the 8 subtrees have their loss rates equal to 1% and the other two have their loss rates equal to 5%. The two subtrees considered by $Al_k(x), |x| = 2$ have the loss rates equal to 1% and 5%, respectively; whereas the two of the three subtrees considered by $Al_k(x), |x| = 3$ have their loss rates equal to 1% and the other has its loss rate equals to 5%. The results are presented in Table III. Compared Table III with Table II, there is no change for the original MLE and there are slight changes for the estimates obtained by $A_k(2)$ and $A_k(3)$. That confirms the robustness of the original MLE and the $A_k(i)$ over $Al_k(x)$. In contrast to the original MLE and $A_k(i)$, the variances of the estimates obtained by $Al_k(x), |x| = 2$ and $Al_k(x), |x| = 3$ have noticeable differences with their counterparts, in particular if the sample size is small because each of the estimators has a descendant with a higher loss rate than that used in the first round. The advantage of this shows at the mean obtained by $Al_k(x), |x| = 2$ that approaches to the true value quicker than that in the first round and that of $Al_k(x), |x| = 3$. This is because one of the two descendants considered by $Al_k(x), |x| = 2$ has a higher loss rate than the other that increases the probability of matching the predicator to its observation. In contrast to $Al_k(x), |x| = 2$, the mean of $Al_k(x), |x| = 3$ has little change from that obtained in the first round. This reflects the tradeoff between efficiency and robustness among $Al_k(x)$, where the larger the $|x|$ is, the robuster the $Al_k(x)$ is to the turbulence of the loss rates in x. To have a similar result as the original MLE, we should select the subtrees that have loss rates equating to 1% for $Al_k(x), |x| = 2$ or 3. Then, the same result as that presented in Table II should be obtained.

To verify the claim made at the end of last paragraph, we conduct another round simulation, where the loss rate of the path of interest is increased from 1% to 5%, and the loss rates of the eight subtrees rooted at node k are divided into two groups, four of them are set to 5% and the other four to 1%. The two estimators from IBE, i.e. $Al_k(x), |x| = 2$ and 3, consider the observations obtained from the subtrees that have their loss rates equal to 1%. The result is presented in Table IV that confirms the estimates of $Al_k(x)$ can be as good as that of the OMLE. In comparison with Table II, there are two noticeable differences in Table IV :

- the means of the estimates approach to the true value quicker; and
- the variances are a magnitude higher.

The first can be derived from Theorem 7 and Corollary 2 since the efficiency of an estimator is inversely proportional to A_k; whereas the second can be obtained from Theorem 8 that states a smaller $\delta_k(x)$ results in a bigger variance.

The simulations show that the original MLE undoultly is the most robust estimator that fits all of the three situations well although it reacts slower than some of the estimators proposed in this paper to the variation of observation. In contrast, there is always an estimator that has a similar performance as that of the MLE in each of the situations. The findings of this paper make it possible to identify a suitable estimator according to end-to-end observation.

VIII. CONCLUSION

This paper starts from finding inspirations that can lead to efficient explicit estimators for loss tomography and ends with a large number of unbiased or asymptotic unbiased and consistent explicit estimators, plus a number of theorems and corollaries to assure the statistical properties of the estimators. One of the most important findings is of the formulae to compute the variances of A_k estimated by the estimators in RSE, IBE and the original MLE. Apart from clearly expressing the connection between the path to be estimated and the subtrees connecting the path to the observers of interest, the formulae potentially have many applications in network tomography, some have been identified in this paper. For instance, using the formulae, we have ranked the MLEs proposed so far, including those proposed in this paper. In addition, the formulae make model selection possible in loss tomography and then the multicast used in end-to-end measurement is no longer only for creating various correlations but also for identifying the subtrees that can be used in estimation. The effectiveness of the strategy has been verified in a simulation study. Apart from those, there are other potentials to use the formulae and the findings that require further exploration.

REFERENCES

[1] Y. Vardi, "Network tomography: Estimating source-destination traffic intensities from link data," *Journal of Amer. Stat. Association*, vol. 91, no. 433, 1996.

[2] R. Cáceres, N. Duffield, J. Horowitz, and D. Towsley, "Multicast-based inference of network-internal loss characteristics," *IEEE Trans. on Information Theory*, vol. 45, 1999.

[3] R. Cáceres, N. Duffield, S. Moon, and D. Towsley, "Inference of Internal Loss Rates in the MBone ," in *IEEE/ISOC Global Internet'99*, 1999.

[4] R. Cáceres, N. Duffield, S. Moon, and D. Towsley, "Inferring link-level performance from end-to-end multicast measurements," tech. rep., University of Massachusetts, 1999.

[5] M. Coates and R. Nowak, "Unicast network tomography using EM algorthms," Tech. Rep. TR-0004, Rice University, September 2000.

[6] B. Xi, G. Nichailidis, and V. Nair, "Estimating netwrok loss rates using active tomography," *Journal of the American Statistical Association*, vol. 101, no. 476, 2006.

[7] T. Bu, N. Duffield, F. Presti, and D. Towsley, "Network tomography on General Topologies," in *SIGCOMM 2002*, 2002.

[8] V. Arya, N. Duffield, and D. Veitch, "Multicast inference of temporal loss charateristics," *Performance Evaluation*, vol. 9-12, 2007.

[9] N. Duffield, J. Horowitz, F. L. Presti, and D. Towsley, "Explicit loss inference in multicast tomography," *IEEE trans. on Information Theory*, vol. 52, no. 8, 2006.

[10] W. Zhu and Z. Geng, "Bottom up inference of loss rate," *Journal of Computer Communications*, vol. 28, no. 4, 2005.

[11] D. Guo and X. Wang, "Bayesian inference of network loss and delay characteristics with applications to tcp performance predication," *IEEE trans. on Signal Processing*, vol. 51, no. 8, 2003.

[12] G. Liang and B. Yu, "Maximum pseudo likelihood estimation in network tomography," *IEEE trans. on Signal Processing*, vol. 51, no. 8, 2003.

[13] Y. Tsang, M. Coates, and R. D. Nowak, "Network delay tomography," *IEEE trans. on Signal Processing*, vol. 51, no. 8, 2003.

[14] F. L. Presti, N. Duffield, J. Horowitz, and D. Towsley, "Multicast-based inference of network-internal delay distributions," *IEEE/ACM trans. on Networking*, vol. 10, 2002.

[15] M.-F. Shih and A. Hero, "Unicast-based inference of network link delay distribution with finite mixture models," *IEEE trans. on Signal Processing*, vol. 51, no. 8, 2003.

[16] E. Lawrence, G. Michailidis, and V. N. Nair, "Network delay tomography using flexicast experiments," *Journal of Royal Statistica Society, Series B*, vol. 68, no. 5, 2006.

[17] K. Harfoush, A. Bestavros, and J. Byers, "Robust identification of shared losses using end-to-end unicast probes," in *Technical Report BUCS-2000-013*, Boston University, 2000.

[18] B. C. Lindsay, "Composite likelihood method," *Contemporary Mathematics*, vol. 80, 1988.

[19] W. Zhu, "An efficient loss rate estimator in multicast tomography and its validity," in *IEEE International Conference on Communicaation and Software*, 2011.

[20] R. Mittelhammer, *Mathematical Statistics for Economics and Business*, vol. 78. Springer, 1996.

[21] J. Besag, "Spatial interaction and the statistical analysis of lattice system (with discussion)," *Journal of Royal Statistical Society*, vol. 36, 1974.

[22] V. M. Joshi, "On the attainment of the cramer-rao lower bound," *Ann. Statist.*, vol. 4, no. 5, pp. 998–1002, 1976.